·医学循证·

基于COSMIN方法对中国腰痛患者罗兰-莫里残疾调查问卷测量性能证据的系统评价

高艺轩[1,2]，王锡友[2]，陈千吉[3]，杨晓明[2]，郭俊明[2]，訾一路[2]，翁志文[2]，马敬祎[2]，章乃文[2]，刘二杨[2]，邵辉[2]，孙亚男[4]，于长禾[2*]

1.100700 北京市，北京中医药大学第一临床医学院
2.100700 北京市，北京中医药大学东直门医院推拿科
3.518001 广东省深圳市罗湖区中医院
4.100053 北京市，首都医科大学宣武医院中医科
*通信作者：于长禾，主治医师；E-mail：yakno2@163.com

【摘要】 **背景** 腰痛全球患病率逐步上升，而腰痛是致残、病休、失业的主要原因，对个人及社会均是沉重的负担。评估慢性腰痛患者的残疾程度对于评价临床干预疗效和临床流行病学至关重要。罗兰-莫里残疾调查问卷（RMDQ）是目前评估腰痛患者残疾的主要工具，但其测量性能在中国人群中的适用性尚不明确。**目的** 评估RMDQ在中国腰痛人群中的适用性，为临床实践、临床研究的应用提供证据。**方法** 计算机检索中国知网、万方数据知识服务平台、SinoMed、PubMed、Embase及Web of Science等，检索时间为建库至2023-10-01，针对腰痛量表的性能研究建立文献库，再从中筛选RMDQ测量性能的研究。按照以共识为基础的健康测量工具的选择标准（COSMIN）系统评价指南对RMDQ量表的测量学性能进行评价，并采用GRADE证据评价等级对证据分级。**结果** 共纳入6篇RMDQ文献，RMDQ内容效度方法学质量为不足，测量性能为充分；内部一致性方法学质量存在非常好与不确定，测量性能为充分；重测性度方法学质量为不确定，测量性能为充分；测量误差方法学质量为不确定，测量性能为充分；效标效度方法学质量为不确定，测量性能为不充分；假设检验方法学质量存在非常好与不确定，测量性能存在充分与不确定；反应度方法学质量存在非常好、足够与不足，测量性能存在充分与不足。根据GRADE证据质量评级结果表明内容效度有低质量证据证明不确定，重测信度及内部一致性有中等质量证据证明充分；测量误差及反应度有低质量证据证明充分；以Oswestry功能障碍指数（ODI）和视觉模拟量表（VAS）为校标时，校标效度有非常低质量证据证明不充分；假设检验有中等质量证据证明不确定。**结论** RMDQ量表方法学质量不高，测量性能尚可，证据质量偏低，在中国腰痛临床实践或试验中需谨慎使用，重测信度与内部一致性虽有中等质量证据证明充分，但研究内容和方法不规范。将来研究应注意规范，能更准确判断在中国人群中的适用性。

【关键词】 腰痛；罗兰-莫里残疾调查问卷；COSMIN；信度；效度；反应度
【中图分类号】 R 681.55 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0519

RMDQ Measurement Performance Evidence in Chinese Patients with Low Back Pain：a Systematic Review Based on COSMIN Method

GAO Yixuan[1,2]，WANG Xiyou[2]，CHEN Qianji[3]，YANG Xiaoming[2]，GUO Junming[2]，ZI Yilu[2]，WENG Zhiwen[2]，MA Jingyi[2]，ZHANG Naiwen[2]，LIU Eryang[2]，SHAO Hui[2]，SUN Yanan[4]，YU Changhe[2]*
1.The first Clinical Department，Beijing University of Chinese Medicine，Beijing 100700，China
2.Tuina and Pain Management Department，Beijing University of Chinese Medicine Affiliated Dongzhimen Hospital，Beijing 100700，China

基金项目：国家自然科学基金青年科学基金项目（81803956）；北京市首发青年优才项目（2020-4-4195）；首都临床特色应用研究（Z181100001718165）

引用本文：高艺轩，王锡友，陈千吉，等.基于COSMIN方法对中国腰痛患者罗兰-莫里残疾调查问卷测量性能证据的系统评价［J］.中国全科医学，2023.［Epub ahead of print］.DOI：10.12114/j.issn.1007-9572.2023.0519.［www.chinagp.net］

GAO Y X，WANG X Y，CHEN Q J，et al. RMDQ measurement performance evidence in Chinese patients with Low back pain：a systematic review based on COSMIN method［J］. Chinese General Practice，2023.［Epub ahead of print］.

3.Shenzhen Luohu District Hospital of Traditional Chinese Medicine，Shenzhen 518001，China
4.Traditional Chinese Medicine Department，Xuanwu Hospital Capital Medical University，Beijing 100053，China
*Corresponding author：YU Changhe，Attending physician；E-mail：yakno2@163.com

【Abstract】 **Background** The global prevalence of low back pain is gradually increasing，and it is the main cause of disability，sick leave，and unemployment，posing a heavy burden on individuals and society. Assessing the degree of disability in patients with chronic low back pain is crucial for evaluating the efficacy of clinical interventions and clinical epidemiology. The Roland Murray Disability Survey Questionnaire（RMDQ）is currently the main tool for evaluating disability in patients with low back pain，but the applicability of its measurement performance in the Chinese population remains unclear. **Objective** To evaluate the applicability of RMDQ in the Chinese population with low back pain and provide evidence for clinical practice and research application. **Methods** CNKI，Wanfang Data Knowledge Service Platform，SinoMed，PubMed，Embase and Web of Science were searched from inception to 2023-10-01，to establish a literature base for the performance of the low back pain scale，and then select research on the measurement performance of RMDQ from it. The measurement performance of the RMDQ scale was evaluated according to the COSMIN system evaluation guidelines，and the evidence evaluation level was used to grade the evidence. **Results** A total of six RMDQ documents were included，with insufficient methodological quality for RMDQ content validity and adequate measurement performance. The quality of internal consistency methodology was very good with uncertainty and measurement performance was adequate； the methodological quality of retesting was uncertain，and the measurement performance was sufficient； the methodological quality of measurement error was uncertain，and the measurement performance was sufficient； the methodological quality of criterion validity was uncertain，and the measurement performance was insufficient； hypothesis testing methodological quality was very good with uncertain，and the measurement performance was sufficient and uncertain； the quality of reactivity methodology was very good，with sufficient and insufficient，while the measurement performance was sufficient with insufficient. According to the GRADE evidence quality rating results，there is low quality evidence to prove uncertainty in content validity，and moderate quality evidence to prove sufficient retesting reliability and internal consistency； there is sufficient evidence of low quality to prove the measurement error and reactivity. There is very low quality evidence of insufficient calibration validity when using the Oswestry Dysfunction Index（ODI）and the Visual Analog Scale（VAS）as calibrators； hypothesis testing had moderate quality evidence of uncertainty. **Conclusion** The methodological quality of the RMDQ scale is not high，with acceptable measurement performanceand low quality of evidence，and needs to be used cautiously in clinical practice or trials of low back pain in China. Although there is sufficient evidence of moderate quality to prove the reliability and internal consistency of retesting，the research content and methods are not standardized. In future research，attention should be paid to standardization to more accurately assess its applicability in the Chinese population.

【Key words】 Low back pain；Roland Murray Disability Survey Questionnaire；COSMIN；Validity；Reliability；Responsiveness

腰痛全球患病率为 13.1%~20.3%[1]，我国患病率为 6.11%~28.5%[2]，腰痛是致残、病休、失业的主要原因[3]，对个人及社会均是沉重的负担。评估慢性腰痛患者的残疾程度对于评价临床干预疗效和临床流行病学至关重要[4]。罗兰－莫里残疾调查问卷（Roland-Morris Disability Questionnaire，RMDQ）是由英国学者 Roland 和 Morris 等设计的一种问卷，是目前评估腰痛患者残疾的主要工具[5]，TSANG 等[6]首次评价 RMDQ 在中国腰痛人群的测量学性能，随后亦有研究，但研究的方法学质量、量表证据强度、是否适用等问题尚不清楚。测量学性能评价决定量表的准确性和可靠度，并影响结果解读。国外对 RMDQ 的以共识为基础的健康测量工具的选择标准（consensus-based standards for the selection of health measurement instruments，COSMIN）研究[7]显示其是高质量证据证明可理解但不全面的问卷，低至非常低质量证据证明其内容效度，中等质量证据证明其结构效度。但其在中国腰痛人群的测量学性能尚不清楚。

COSMIN 是通过评估患者报告结局指标测量工具研究的方法学质量及测量工具测量学性能而制定的，目的为指导临床研究使用适当的测量工具，提高研究质量[8]。COSMIN 系统评价指南[9]对测量工具测量学性能的系统评价进行了详细阐述，并指导了研究步骤。

本研究按照 COSMIN 系统评价指南对 RMDQ 量表的测量学性能进行评价，并采用 GRADE 证据评价等级对证据分级，目的是评估其在中国腰痛人群中的适用性，为临床实践、临床研究的应用提供证据。

1 资料和方法

1.1 测量工具

RMDQ 是从疾病影响量表（Sickness Impact Profile，SIP）中选择了 24 个受腰痛特异影响的问题组成，旨在评估腰痛患者的功能状态[5]。每个问题以“因为我的腰痛”加以限制，与其他导致功能障碍的原因区分，方便患者回答。24 个问题主要涉及腰痛对坐、卧、走、自理能力与日常生活等方面的影响，每题 1 分，“是”为 1 分，“否”为 0 分，总分 0~24 分，得分越高，功能障碍程度越高[5]。

1.2 检索策略与数据库建立

针对腰痛量表的性能研究建立文献库，再从中筛选 RMDQ 测量性能的研究。从数据库创建至 2023-10-01，计算机检索中国知网、万方数据知识服务平台、SinoMed、PubMed、Embase 及 Web of Science 等，中文关键词包括：（1）腰疼、腰痛、腰背痛、腿痛等；（2）中国；（3）效度、信度、反应度等测量学性能。英文关键词包括：（1）back pain、low back pain；（2）China、Chinese、Hong Kong；（3）inter method comparison、data collection method、validation study、outcome measure；（4）internal consistency、reliab、valid、retest。

1.3 文献纳入与排除标准

纳入标准：（1）中国腰痛人群临床测量（包括诊断、评估和预测等）研究；（2）包括 RMDQ 量表 9 种测量性能（内容效度、结构效度、假设检验、内部一致性、重测信度、测量误差、标准效度、反应度及跨文化效度）至少 1 种研究；（3）全文可用。排除将 RMDQ 用于另一种工具测量性能研究（如选择 RMDQ 作为对比量表以评价 PSEQ 结构效度的研究）。

1.4 文献筛选与数据提取

2 位研究人员独立排除无关研究，再检索保留文献全文内容，按照纳入及排除标准对全文数据提取，结果如有不同，双方进行协商，若仍不同则由第三方仲裁。每项研究应提取样本量、性别、年龄、地区、语言、疾病特征、填写 RMDQ 的场景以及测量性能的研究结果与统计学方法等。

1.5 评价流程

COSMIN 对量表的系统评价分为 3 个部分、10 个步骤（图 1）：第一部分文献检索，步骤 1~4，内容为明确评价目的、制定纳排标准、实施文献检索、获取与筛选。第二部分评价测量性能，步骤 5~7，内容为评价内容效度、内部结构和其他测量性能。其中包含对研究偏倚风险、测量性能优劣的评价及结果汇总，后运用 GRADE 系统形成推荐等级。第三部分选择量表，步骤 8~10，内容是描述可解释性和可行性、形成量表推荐意见。

1.5.1 研究的方法学质量评价：按照“COSMIN 偏倚风险清单”[10]进行评价，结果有“非常好”“足够”“不确定”或“不足”。每项研究质量的总分按框中条目的最低分给出。如果多种测量性能在一项研究中，则每种测量性能均需进行评价。

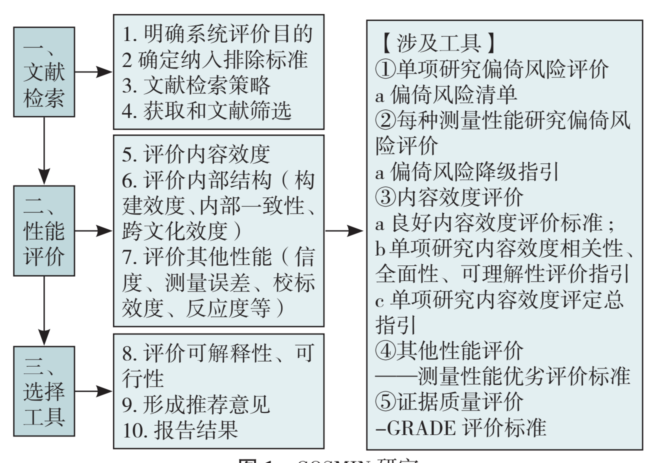


图 1 COSMIN 研究

Figure 1 COSMIN Study

1.5.2 测量性能的评价顺序：首先评价量表开发质量及内容效度。内容效度在测量性能中至关重要[11]，先排除内容效度不足的研究。后评价内部结构，如内部一致性、结构效度等。最后考虑其他测量性能，如重测信度、反应度等。

1.5.3 测量性能的评价：内容效度评价按照“内容效度 10 条评价标准”[11]，评价结果为充分（+）、不足（–）或不确定（？）。根据“单个研究内容效度评级指南及评定指南”进行评价，结果为充分（+）、不足（–）、不一致（±）或不确定（？）。其他测量性能根据“更新测量性能优劣评价标准”[12-13]将测量性能评价为充分（+）、不足（–）或不确定（？）。

1.5.4 证据综合与证据质量分级：评判每个测量特性可用研究结果的一致性。结果一致时，可对结果进行定量汇总或定性总结，并与测量特性的良好标准进行比较，来确定量表的测量特性总体上是充分（+）、不足（–）、不一致（±）还是不确定（？）；如果结果不一致，可用亚组分析或根据大多数一致的结果得出结论，并按照不一致性降低证据质量。

COSMIN 应用“改进 GRADE 证据质量评级”标准[14]，标准包含：偏倚风险[15]、不一致性、不精确性及间接性。研究质量可根据以上标准进行降级。

2 结果

2.1 纳入研究一般情况

共检索到 4 731 篇文献，1 394 篇文献因重复被删除，3 179 篇文献因阅读题目摘要后不相关被删除，最后 158 篇文献被纳入数据库。全文浏览后，11 篇 RMDQ

量表测量性能研究被纳入，其中 2 篇无法获得全文与 3 篇重复的研究（2 篇毕业论文与 2 篇中文文献为同一研究，1 篇英文文献与 1 篇中文文献为相同研究）被排除在外，最终纳入文献 6 篇[4-6，16-18]。文献检索流程图见图 2，纳入文献基本情况见表 1。

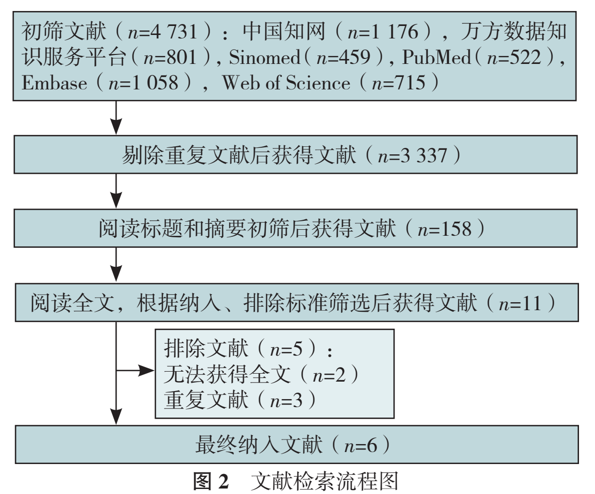


图 2 文献检索流程图
Figure 2 Literature retrieval flow chart

表 1 纳入文献基本情况
Table 1 Basic information of the included literature

第一作者	发表时间(年)	RMDQ 版本	总样本量(例)	平均年龄(岁)	女性(%)	疾病特点
YAO[4]	2020	中文版	353	50.53	55.81	LDH
FAN[16]	2013	简体中文版	116	52.17	40.51	LBP
YI[17]	2012	简体中文版	187	41.30	49.73	LBP
高明暄[18]	2005	中文版	57	43.6	40.35	LBP
何高[5]	2005	中文版	48	38.5	37.50	LBP
TSANG[6]	2004	香港中文版	112	41.6	57.10	LBP[a]

注：LDH= 腰椎间盘突出症，LBP= 腰痛；[a] 包括坐骨神经痛。

2.2 纳入研究的方法学质量评价

2.2.1 内容效度：何高[5]研究中仅提及采用专家审定方法对内容效度进行评价，缺少必要的方法学描述。TSANG[6]对各条目认可比例、区分度和修正项的合计相关系数来评价内容效度。根据 COSMIN 标准，需有相关领域专家评价及数据的处理和方法，两项研究各有欠缺，故评为不足。

2.2.2 内部一致性：YAO 等[4]、高明暄等[18]及 YI 等[17]研究仅分析了总量表的克隆巴赫系数，缺少单独条目的分析，故均评为不确定。何高[5]与 FAN 等[16]分析了总量表及单独条目的克隆巴赫系数，评为非常好。TSANG[6]将 KR20 系数大小作为对内部一致性的评估，KR20 系数被看作是克隆巴赫系数的特殊形式，因此评为非常好。

2.2.3 重测信度：YAO 等[4]、何高[5]、TSANG[6]、高明暄等[18]均未说明 2 次填写量表的环境，无法判断两次填写条件的相似性；FAN 等[16]缺少对纳入患者干预的描述，无法判断两次填写时间间隔的合适性及期间内待测结构是否稳定；YI 等[17]患者 2 次填写时间间隔为 24 h，考虑间隔时间不合适，且未说明 2 次填写的环境，以上研究均评为不确定。

2.2.4 测量误差：YAO 等[4]研究未描述两次量表填写的条件，无法评价，故评为不确定。

2.2.5 校标效度：何高[5]研究以 Oswestry 功能障碍指数(Oswestry disability index，ODI)和视觉模拟量表(Visual Analogue Scale，VAS)为校标，采用 Spearman 相关系数对 294 例患者进行评估，但目前没有统一的腰痛金标准量表，因此研究质量评为不确定。

2.2.6 假设检验：YAO 等[4]分别以日本骨科协会腰痛评分量表（Japanese Orthopaedic Association，JOA）5 个问题、数字疼痛评定量表（Numerical Pain Rating Scale，NPRS）、ODI、36 项健康调查简表（Short Form 36 Health Survey，SF-36）为对比工具评价聚合效度，但未提供对比工具测量性能相关证据被评为不确定。何高[5]提出两种理论假设，使用配对性 t 检验来测量样本，假设与理论一致，评为非常好。

2.2.7 反应度：YAO 等[4]、TSANG[6]均采用曲线下面积(area under the curve，AUC)评估反应度，YAO 等[4]通过亚组间数据统计分析评价，描述了亚组重要特征、统计方法合适且研究设计无重要缺陷，评为非常好。TSANG[6]未具体描述使用的标准，评为足够。何高[5]用标准化反应均数（standardized response mean，SRM）评估反应度，未充分描述统计方法，评为足够。

研究缺少对 RMDQ 跨文化和结构效度的评价（表 2）。

2.3 量表测量学性能及其评价

2.3.1 内容效度：何高[5]研究包含了与腰痛密切的内容，专家评审为内容效度良好。TSANG[6]研究除个别条目，RMDQ 项目的认可比例均在 0.20~0.80；条目区分度指数为 0.30~0.73；条目的二序相关点为 0.31~0.68；故内容效度良好。二者研究均可看作专业人员对其内容效度进行评价，但评价方法不符合 COSMIN 标准，评为不确定。

2.3.2 内部一致性：5 项研究[4-5，16-18]克朗巴赫系数均 >0.7，分别为 0.807、0.84、（城市 0.874、农村 0.883）、0.885、0.826；TSANG[6] KR20 系数为 0.86，说明内部一致性皆良好，评为充分。

2.3.3 重测信度：6 项研究[4-6，16-18]组内相关系数(intraclass correlation coefficient，ICC)结果分别为 0.855、0.947、（城市组 0.952、农村组 0.949）、0.95、0.945、0.91，均 >0.7，评为充分。

2.3.4 测量误差：YAO 等[4]计算标准测量误差(standard

error of mean，SEM）结果为0.28、最小可检测差异（smallest detectable change，SDC）结果为0.77、最小临床重要变化（minimum important change，MIC）结果为1.74，根据SDC小于MIC，评为充分。

2.3.5 校标效度：何高[5]以ODI和VAS为校标，Spearman相关系数为0.3~0.5，<0.7，故评为不充分。

2.3.6 假设检验：YAO等[4]研究中，RMDQ与JOA问题1~5、NPRS、ODI、SF-36的相关系数为-0.76~0.74，研究中未提出具体假设或相关标准描述，故评为不确定。何高[5]研究基于两种理论假设：1周内打算或已看过医生的腰痛患者其功能障碍情况应该比没有看过或打算来看医生的患者的功能障碍情况更重；1周内需要吃药的腰痛患者其功能障碍情况应该比没有吃过药的患者的功能障碍情况更重。研究配对 t 检验结果显示1周内接受看医或吃药的患者的功能障碍情况更重。假设与理论一致，评为充分。

2.3.7 反应度：YAO等[4]AUC结果为0.868，评为充分。何高[5]用SRM评估反应度，SRM值为0.7~0.9，反应度较好，评为充分。TSANG[6]AUC结果较为多变，不同标准下结果为0.67~0.84，评为不充分。

无研究评价RMQD量表的跨文化、结构效度（表2）。

表2 纳入研究方法学质量评价与测量性能评价结果

Table 2 **Results** of methodological quality evaluation and measurement performance evaluation of the included studies

测量性能	第一作者	方法学质量评价	测量性能评价
内容效度	何高[5]	不足	+
	TSANG[6]	不足	+
内部一致性	YAO[4]	不确定	+
	FAN[16]	非常好	+
	YI[17]	不确定	+
	高明暄[18]	不确定	+
	何高[5]	非常好	+
	TSANG[6]	非常好	+
重测信度	YAO[4]	不确定	+
	FAN[16]	不确定	+
	YI[17]	不确定	+
	高明暄[18]	不确定	+
	何高[5]	不确定	+
	TSANG[6]	不确定	+
测量误差	YAO[4]	不确定	+
校标效度	何高[5]	不确定	-
假设检验	YAO[4]	不确定（与JOA问题1~5、NPRS、ODI、SF-36比较）	?
	何高[5]	非常好	+
反应度	YAO[4]	非常好	+
	何高[5]	足够	+
	TSANG[6]	不足	-

注：+表示测量性能充分；-表示测量性能不足；？表示测量性能不确定；JOA=日本骨科协会腰痛评分量表，NPRS=数字疼痛评定量表，ODI=Oswestry功能障碍指数，SF-36=36项健康调查简表。

2.4 证据综合与证据质量分级

研究采用GRADE证据评价等级对证据进行分级。YAO等[4]、高明暄等[18]、何高[5]中文版与FAN等[16]、YI等[17]简体中文版及TSANG[6]香港中文版的量表内容差异不大，故综合了这6篇研究证据。

GRADE证据综合与证据质量评级结果显示：RMDQ内容效度不确定，因偏倚风险严重降级为低质量；结构效度无相关研究；内部一致性充分，证据等级因偏倚风险严重降级为中等质量；重测信度充分，证据等级因偏倚风险严重降级为中等质量；测量误差充分，证据等级因偏倚风险非常严重降级为低质量；以VAS为校标的研究，其结果校标效度不充分，证据等级因偏倚风险非常严重及结果不精确严重降级为非常低质量；假设检验不确定，证据等级因偏倚风险严重降级为中等质量；反应度充分，证据等级因偏倚风险非常严重降级为低质量。所有研究证据在不一致性、不准确性及间接性三个方面良好，均不存在降级情况（表3）。

表3 GRADE证据综合与证据质量评级结果

Table 3 GRADE evidence summary and evidence quality rating results

测量属性	文献数量（篇）	证据质量	证据综合	证据	证据质量降低原因
内容效度	2[5-6]	低质量	不确定	++--	偏倚风险严重
内部一致性	6[4-6, 16-18]	中等质量	充分	+++-	偏倚风险严重
重测信度	6[4-6, 16-18]	中等质量	充分	+++-	偏倚风险严重
测量误差	1[4]	低质量	充分	++--	偏倚风险非常严重
校标效度	1[5]	非常低质量	不充分	+---	偏倚风险非常严重、结果不精确严重
假设检验	2[4-5]	中等质量	不确定	+++-	偏倚风险严重
反应度	3[4-6]	低质量	充分	++--	偏倚风险非常严重

3 讨论

本研究采用COSMIN标准对RMDQ在中国人群中方法学质量与测量性能进行评价。综合评价结果来看，内容效度和反应度呈低质量，且缺少结构效度证据，尽管信度的测量学性能较为充分，但目前研究存在严重偏倚风险，RMDQ临床应用的结果和解读需要谨慎。

内容效度是最重要的测量学性能。条目是量表、反映预期概念的基本单元，其内容和质量直接影响量表的内容效度。因此，COSMIN首先评价量表的内容效度，包括相关人群（专家与受试者）的定性或定量研究、对量表相关性、全面性及可理解性全方位评价。本研究中两篇文献一是缺少数据的处理和方法，一是未说明进行评价的是否为专家，故两项研究的质量评价为不足。

效标效度是对结构效度和内容效度的补充，当“金标准”有异议时，校标效度和聚合效度将变得难以区分。

本研究中效标效度的证据质量非常低，测量性能不充分，主要表现为与 ODI 和 VAS 校标的相关性低，降低了证据质量。目前并没有统一的腰痛金标准量表，ODI 是功能障碍量表，病情程度相对比 RMDQ 要轻，若作为效标或金标准是不完全合适的；VAS 仅评价了疼痛，而 RMDQ 不仅针对疼痛，若作为效标亦是不合适的。

假设检验是先对总体特征作出某种假设，通过抽样研究的统计推理，对此假设应该被拒绝还是接受作出推断，来验证总体的特征。在研究中应明确给出假设，选取与其测量内涵相同 / 相似或不相同的量表，说明两者的关系是正 / 负相关性及相关程度等，而且选择的量表应该本身具有较为充分的测量学性能，并提供相关证据。YAO 等[4]仅提出 RMDQ 与 JOA、NPRS、ODI、SF-36 具有相关性的假设，但缺少具体假设或相关标准描述，评价为中等质量的不确定。何高[5]研究给出了明确假设，通过检验证实假设与理论一致，评为充分。此项测量性能存在非常好与不确定两种情况，存在偏倚，被评价为中等质量。

信度是反映测验结果受到随机误差影响程度的指标，其与效度的关系是必要但不充分。RMDQ 信度证据均为中等质量，其中内部一致性为中等质量，测量性能充分，方法学部分因其未计算每一条目的内部一致性系数从而存在偏倚，降低了证据质量；重测信度亦为中等质量，测量性能充分，但无法判断两次填写条件的相似性、两次填写时间间隔的合适性及在这期间内待测结构是否稳定，使得证据质量降低。在以后的研究中应注意重测间隔时间、地点及其他条件是否合适，以提高重测信度的质量。

反应度指量表反映微小特性变化的能力，是对效度进一步的评价。反应度根据变化程度和是否有临床意义分为 SDC 和 MIC，前者只要求量表最小评分发生变化，无论临床是否有意义；后者强调量表最小评分变化要有临床意义，量表的可解释性属于后者。RMDQ 反应度被判定为低质量，测量性能有充分与不充分两种不一致的结果，方法学存在差异，使结果发生偏倚降低了证据质量。同时也提示 RMDQ 量表在反映患者治疗效果上存在着不足。

整体来看，RMDQ 量表的信度比效度具有高证据质量。信度是反应测量结果的一致性和稳定性程度，效度是反映实际能测出的有效程度。信度是效度的必要但不充分条件，无信度必然无效度，但有信度未必有效度；效度是信度的充分但不必要条件，有效度必然会有信度，但无效度却未必无信度。RMDQ 效度证据质量低或缺失，说明其在中国腰痛人群中的适用性有待提高。尽管 RMDQ 为国际公认通用量表，但并不能说明其可以很好应用于中国临床。研究也指出 RMDQ 量表是一个可理解但不全面的量表，其内容效度证据呈低至极低质量[7]，本研究结果与之相同。

研究结果显示方法学质量普遍较低，统计方法与研究设计选择不当、分析不足、样本量不够等会降低质量。纳入 6 篇文献共涉及 3 个 RMDQ 版本，且为 RMDQ 英文翻译版，但缺少 RMDQ 跨文化效度研究，那么对于不同版本便有不同的翻译方式，可能会使 RMDQ 量表在某些问题上产生理解偏差从而导致结果发生偏倚。同时也提示当前没有统一的中文版 RMDQ 量表。跨文化效度研究需在文化调适的基础上再进行结构效度的评价，评价时应选用两组样本，除了语言的差异外，应控制其余人群特征（如年龄、性别、疾病特点等）的一致。

本研究也存在着很多不足之处，纳入文献使用的量表没有全部获取，纳入评价的文献数量较少，也没有比较和分析各版本间的差异，从而具有局限性。COSMIN 系统评价指南于 2005 年发布，但研究纳入文献有 3 篇发表于 2010 年或更早，其研究方法学质量按照较新标准进行评价存在争议，因此研究评价的结果可能存在偏倚。在此后的研究中应当获取所有纳入文献的量表，对其进行比较；此外应尽量纳入发布于 2010 年之后的文献，以减少争议。量表性能测量的研究设计、研究方法等应更标准、详细。还应结合中国实际情况，改进量表，得到适用于中国本土腰痛患者评估残疾的量表。测量学性能是一个不断累积的过程，而且随着文化、语言、年代、测量方式、人群认识概念、目标人群特点或者疾病谱等的变化而不断增加新的证据[19]；FDA 也指出量表的研制和修改是一个不断增加证据、非线性轮状的过程[20]。

4 结论

RMDQ 量表方法学质量不高，测量性能尚可，证据质量偏低，在中国腰痛临床实践或试验中需谨慎使用，重测信度与内部一致性虽有中等质量证据证明充分，但研究内容和方法不规范。将来研究应注意规范，能更准确判断在中国人群中的适用性。

作者贡献：高艺轩进行文章撰写，文献筛选；王锡友、孙亚男提供指导性支持；陈千吉进行文献筛选；杨晓明、郭俊明、刘二杨、邵辉负责文章内容、质量评价校对；訾一路、翁志文、马敬祎、章乃文负责方法学质量评价；于长禾负责方法学设计与文章质量监控，对论文负责。

本文无利益冲突。

参考文献

[1] MEUCCI R D，FASSA A G，FARIA N M X. Prevalence of chronic low back pain：systematic review［J］. Rev Saude Publica，2015，49：1. DOI：10.1590/S0034-8910.2015049005874.

[2] 陈栋，陈春慧，胡志超，等. 中国成人腰痛流行病学的系统评

价［J］. 中国循证医学杂志，2019，19（6）：651-655. DOI：10.7507/1672-2531.201801044.

［3］GBD Disease and Injury Incidence and Prevalence Collaborators. Global，regional，and national incidence，prevalence，and years lived with disability for 328 diseases and injuries for 195 countries，1990-2016：a systematic analysis for the Global Burden of Disease Study 2016［J］. Lancet，2017，390（10100）：1211-1259. DOI：10.1016/S0140-6736（17）32154-2.

［4］YAO M，XU B P，LI Z J，et al. A comparison between the low back pain scales for patients with lumbar disc herniation：validity，reliability，and responsiveness［J］. Health Qual Life Outcomes，2020，18（1）：175. DOI：10.1186/s12955-020-01403-2.

［5］何高 . 中文版 RMDQ 下腰痛量表的研制［D］. 合肥：安徽医科大学，2005.

［6］TSANG R C C. Measurement properties of the Hong Kong Chinese version of the roland-morris disability questionnaire［J］. Hong Kong Physiother J，2004，22（1）：40-49. DOI：10.1016/S1013-7025（09）70049-1.

［7］CHIAROTTO A，OSTELO R W，BOERS M，et al. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain［J］. J Clin Epidemiol，2018，95：73-93. DOI：10.1016/j.jclinepi.2017.11.005.

［8］MOKKINK L B，PRINSEN C A，BOUTER L M，et al. The COnsensus-based Standards for the selection of health Measurement INstruments（COSMIN）and how to select an outcome measurement instrument［J］. Braz J Phys Ther，2016，20（2）：105-113. DOI：10.1590/bjpt-rbf.2014.0143.

［9］PRINSEN C A C，MOKKINK L B，BOUTER L M，et al. COSMIN guideline for systematic reviews of patient-reported outcome measures［J］. Qual Life Res，2018，27（5）：1147-1157. DOI：10.1007/s11136-018-1798-3.

［10］MOKKINK L B，DE VET H C W，PRINSEN C A C，et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures［J］. Qual Life Res，2018，27（5）：1171-1179. DOI：10.1007/s11136-017-1765-4.

［11］TERWEE C B，PRINSEN C A C，CHIAROTTO A，et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures：a Delphi study［J］. Qual Life Res，2018，27（5）：1159-1170. DOI：10.1007/s11136-018-1829-0.

［12］PRINSEN C A C，VOHRA S，ROSE M R，et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline［J］. Trials，2016，17（1）：1-10. DOI：10.1186/s13063-016-1555-2.

［13］TERWEE C B，BOT S D M，DE BOER M R，et al. Quality criteria were proposed for measurement properties of health status questionnaires［J］. J Clin Epidemiol，2007，60（1）：34-42. DOI：10.1016/j.jclinepi.2006.03.012.

［14］GRADE. GRADE Handbook-Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach［EB/OL］.（2013）［2023-03-13］. https://gdt.gradepro.org/app/handbook/handbook.html.

［15］TERWEE C B，MOKKINK L B，KNOL D L，et al. Rating the methodological quality in systematic reviews of studies on measurement properties：a scoring system for the COSMIN checklist［J］. Qual Life Res，2012，21（4）：651-657. DOI：10.1007/s11136-011-9960-1.

［16］FAN S W，HU Z A，HONG H，et al. Cross-cultural adaptation and validation of simplified Chinese version of the Roland-Morris Disability Questionnaire［J］. Spine，2012，37（10）：875-880. DOI：10.1097/BRS.0b013e31823b0460.

［17］YI H L，JI X R，WEI X Z，et al. Reliability and validity of simplified Chinese version of Roland-Morris questionnaire in evaluating rural and urban patients with low back pain［J］. PLoS One，2012，7（1）：e30807. DOI：10.1371/journal.pone.0030807.

［18］高明暄，刘兴炎 . 中文版 Roland-Morris 腰痛失能问卷可靠性评定［J］. 实用医学杂志，2005，21（24）：2755-2756. DOI：10.3969/j.issn.1006-5725.2005.24.015.

［19］REEVE B B，WYRWICH K W，WU A W，et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research［J］. Qual Life Res，2013，22（8）：1889-1905. DOI：10.1007/s11136-012-0344-y.

［20］Department of Health and Human Services FDA Center for Drug Evaluation and Research U S，Department of Health and Human Services FDA Center for Biologics Evaluation and Research U S，Department of Health and Human Services FDA Center for Devices and Radiological Health U S. Guidance for industry：patient-reported outcome measures：use in medical product development to support labeling claims：draft guidance［J］. Health Qual Life Outcomes，2006，4：79. DOI：10.1186/1477-7525-4-79.

（收稿日期：2023-07-14；修回日期：2023-11-28）

（本文编辑：贾萌萌）